STOCHASTIC MODELING OF BIOCHEMICAL REACTIONS

Abhyudai Singh and João Pedro Hespanha*
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93101.

Abstract

The most common theoretical approach to model the interactions in a biochemical process is through chemical reactions. Often for these reactions, the dynamics of the first **M**-order statistical moments of the species populations do not form a closed system of differential equations, in the sense that the time-derivatives of first **M**-order moments generally depend on moments of order higher than **M**. However, for analysis purposes, these dynamics are often made to be closed by approximating the needed derivatives of the first **M**-order moments by nonlinear functions of the same moments. These functions are called the *moment closure functions.*

This paper presents a systematic procedure to construct these moment closure functions. This is done by first assuming that they exhibit a certain *separable* form, and then matching time derivatives of the exact (not closed) moment equations with that of the approximate (closed) equations for some initial time and set of initial conditions. Using these results a stochastic model for gene expression is investigated. We show that in gene expression mechanisms, in which a protein inhibits its own transcription, the resulting negative feedback reduces stochastic variations in the protein populations.

1 Introduction

This paper presents a systematic procedure for constructing approximate stochastic models for chemical reactions used for modeling biochemical processes such as gene regulatory networks. Such models are motivated by the recent work [1, 2] which provide considerable experimental evidence for stochastic fluctuations in gene expression and regulation and may account for the large amounts of cell to cell variation observed in genetically identical cells exposed to the same environment conditions [3, 4]. Furthermore, studies of engineered genetic circuits designed to act as toggle switches or oscillators have revealed large stochastic effects. Stochastically is therefore an inherent feature of biological dynamics and developing stochastic models which capture this stochastically have become increasingly important. Analysis of such model not only helps us to discover the benefits that biology draws from stochasticity but also helps to explore its use for other applications, for example, design of stochastic bio-inspired decision algorithms to control the motion of networks of artificial mobile agents such as small autonomous micro-UAVs or UGVs involved in surveillance and/or tracking missions. Although our work focuses on biochemical reactions, the modeling tools developed in this paper can be applied to a very general class of stochastic systems, in particular Markovian systems whose state evolves due to events triggered by stochastic processes. Other examples of such systems include ecological systems where these events correspond to births, recruitment or deaths and in networked control systems where the events correspond to exchanges of messages, communication faults, etc.

The most common theoretical approach is to model the interactions in a biochemical process as chemical reactions. The stochasticity in the biochemical process is then captured using the stochastic formulation of chemical kinetics which treats the reactions as probabilistic events. The time evolution of the system is then described by a single equation for a probability density function, where time and species populations appear as independent variables, called the Chemical Master Equation (CME) [5, 6]. This equation can only be solved analytically for relatively few, highly idealized

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Stochastic Modeling Of Biochemical Reactions	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Electrical and Computer Engineering University of California, Santa Barbara, CA 93101**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**7**	

cases and generally Monte Carlo simulation techniques are used which are also known as the Stochastic Simulation Algorithm (SSA) to study stochasticity in bio-chemical reactions [7, 8, 4, 9, 10]. Since one is often interested in only the first and second order statistical moments for the number of molecules of the different species involved, much effort can be saved by applying approximate methods to produce these low-order moments, without actually having to solve for the probability density function. Various such approximate methods have been developed, for example, using the Fokker-Plank approximation, expanding the Master equation, etc [5, 11]. In this paper, an alternative approximate method for estimating lower-order moments is introduced using moment closure techniques.

We show that the biochemical reactions can be conveniently modeled as a Stochastic Hybrid System (SHS), the state of which are the populations of different species involved in the reactions [12]. Then, the time evolution of the moments of the population is obtained using results from the SHS literature. It has been shown that for reactions with more than one reactants, time evolution of the first **M** order moments of the population is not closed, in the sense that it depends on moments of order higher than **M**. For analysis purpose, the time evolution of the first **M** order moments is made to be closed by approximating these higher order moments as a nonlinear function of moments up to order **M**, which we refer to as the *moment closure function*. This paper presents explicit formulas to construct these moment closure functions using the recently introduced techniques [13, 14] of matching time derivatives between the the exact (not closed) moment equations with that of the approximate (closed) equations for some initial time and set of initial conditions. The striking feature of these formulas are that they are independent of the reaction parameters (reaction rates and stoichiometry), moreover, the accuracy of the approximation can be improved by increasing **M**.

These closed moment equations provide time evolution of lower order moments for populations of species involved in the biochemical reactions. Apart from providing fast simulation times and lesser computation burden compared to Monte Carlo simulations these approximate models also open the doors to other types of analysis tools, for example, sensitivity analysis of chemical master equation with respect to reaction parameters. However, they provide lesser information about the probability distribution as compared to Monte Carlo simulations. To illustrate the applicability of our results, we investigate a stochastic model for gene expression. We show that a negative feedback mechanism caused by the expressed protein inhibiting its own transcription, reduces stochastic variations in the protein populations.

2 Stochastic Modeling of Chemically Reacting Systems

Consider a system of n species X_j, $j \in \{1,\dots,n\}$ inside a fixed volume V involved in K reactions of the form

$$R_i : u_{i1}X_1 + \ldots + u_{in}X_n \quad \xrightarrow{c_i} \quad v_{i1}X_1 + \ldots + v_{in}X_n + * \tag{1}$$

for all $i \in \{1,\dots,K\}$, where $u_{ij} \in \mathbb{N}_{\geq 0}$ is the stoichiometry associated with the j^{th} reactant in the i^{th} reaction and $v_{ij} \in \mathbb{N}_{\geq 0}$ is the stoichiometry associated with the j^{th} product in the i^{th} reaction, and $*$ represents products other than the species X_j. As all chemical reactions occur in a series of elementary reactions [15], which are generally uni- or bi-molecular, we assume

$$u_{i1} + \ldots + u_{in} \leq 2, \ \ \forall i \in \{1,\dots,K\}, \tag{2}$$

and hence, we only allow reactions which have the form given in the first column of Table 1. The reaction parameter c_i characterizes the reaction R_i and, together with the stoichiometry, defines the probability that a particular reaction takes place in an "infinitesimal" time interval $(t, t+dt]$. This probability is given by the product $c_i h_i dt$ where h_i is the number of distinct molecular reactant combinations present in V at time t for the reaction R_i and $c_i dt$ is the probability that a particular reactant combination of R_i will actually react on $(t, t+dt]$. The number, h_i depends both on the reactants stoichiometry u_{ij} in R_i and on the number of reactant molecules in V. Table 1 shows the value of h_i for different reaction types [7]. In this table and in the sequel, we denote by $\mathbf{x}_j$, the number of molecules of the species X_j in the volume V and $\mathbf{x} := [\mathbf{x}_1,\dots,\mathbf{x}_n]^T \in \mathbb{R}^n$.

To model the time evolution of the number of molecules $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, a special class of Stochastic Hybrid Systems (SHS) were introduced in [12] . More specifically, to fit the framework of our problem, these system are characterized by trivial dynamics

$$\dot{\mathbf{x}} = 0, \ \ \mathbf{x} = [\mathbf{x}_1,\dots,\mathbf{x}_n]^T, \tag{3}$$

Table 1: $h_i(\mathbf{x})$ for different reaction types.

Reaction R_i	$h_i(\mathbf{x})$
$X_j \longrightarrow *$	$\mathbf{x}_j$
$X_j + X_t \longrightarrow *, \ (t \neq j)$	$\mathbf{x}_j \mathbf{x}_t$
$2X_j \longrightarrow *$	$\frac{1}{2}\mathbf{x}_j(\mathbf{x}_j - 1)$

a family of *K reset maps*

$$\mathbf{x} = \phi_i(\mathbf{x}^-), \ \ \phi_i : \mathbb{R}^n \to \mathbb{R}^n, \tag{4}$$

and a corresponding family of *K transition intensities*

$$\lambda_i(\mathbf{x}), \ \ \lambda_i : \mathbb{R}^n \to [0, \infty) \tag{5}$$

for all $i \in \{1, \ldots, K\}$. Each of the reset maps $\phi_i(\mathbf{x})$, and corresponding transition intensities $\lambda_i(\mathbf{x})$ are uniquely defined by the i^{th} reaction and given by

$$\mathbf{x} \mapsto \phi_i(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 - u_{i1} + v_{i1} \\ \mathbf{x}_2 - u_{i2} + v_{i2} \\ \vdots \\ \mathbf{x}_n - u_{in} + v_{in} \end{bmatrix}, \ \lambda_i(\mathbf{x}) = c_i h_i(\mathbf{x}) \tag{6}$$

for all $i \in \{1, \ldots, K\}$. In essence, if no reaction takes place, the state remains constant and whenever the i^{th} reaction takes place, $\phi_i(\mathbf{x})$ is "activated" and the state $\mathbf{x}$ is reset according to (6), furthermore, the probability of the activation taking place in an "infinitesimal" time interval $(t, t + dt]$ is $\lambda_i(\mathbf{x})dt$.

3 Moment Dynamics

Given a vector $\mathbf{m} = (m_1, m_2, \ldots, m_n) \in \mathbb{N}^n_{\geq 0}$ of n greater than equal to zero integers, we define the *(uncentered)* moment of $\mathbf{x}$ associated with $\mathbf{m}$ to be

$$\mu^{(\mathbf{m})}(t) = \mathbf{E}\left[\mathbf{x}^{(\mathbf{m})}(t)\right], \qquad \forall t \geq 0 \tag{7}$$

where $\mathbf{E}$ stands for the expected value and

$$\mathbf{x}^{(\mathbf{m})} := \mathbf{x}_1^{m_1} \mathbf{x}_2^{m_2} \cdots \mathbf{x}_n^{m_n}. \tag{8}$$

The sum $\sum_{j=1}^n m_j$ is called the *order of the moment* $\mathbf{m}$.

Using results from the SHS literature, more specifically by applying Theorem 1 in [16] to the SHS (3)-(5), one can show that the time derivative of a moment $\mu^{(\mathbf{m})}$ is

$$\dot{\mu}^{(\mathbf{m})} = \mathbf{E}\left[\sum_{i=1}^{K} \left(\phi_i(\mathbf{x})^{(\mathbf{m})} - \mathbf{x}^{(\mathbf{m})}\right) \lambda_i(\mathbf{x})\right]. \tag{9}$$

If any of the reactions have more than one reactants, i.e, has the form $X_j + X_t \xrightarrow{c_i} *$ for which $\lambda_i(\mathbf{x})$ is quadratic in $\mathbf{x}$, then, from (9) it can been shown that the time derivative of a moment of order m^*, is given by a linear combination of moments of orders upto $m^* + 1$ [17]. Hence, if one stacks all moments in an infinite vector

$$\mu_\infty = [\mu^{(\mathbf{m}1)}, \mu^{(\mathbf{m}2)}, \cdots]^T, \ \ \mathbf{m}p \in \mathbb{N}^n_{\geq 0}, \ \ \forall p \in \{1, 2, \ldots\} \tag{10}$$

its dynamics can be written as

$$\dot{\mu}_\infty = A_\infty \mu_\infty, \tag{11}$$

for some infinite matrix A_∞. As the above infinite dimensional system cannot be solved analytically, we truncate (11) by creating a vector

$$\mu = [\mu^{(\mathbf{m}1)}, \mu^{(\mathbf{m}2)}, \ldots, \mu^{(\mathbf{m}k)}]^T \in \mathbb{R}^k \tag{12}$$

containing the top k elements of μ_∞ which correspond to the lower-order moments of interest. Then, (11) can be rewritten as

$$\dot{\mu} = I_{k \times \infty} A_\infty \mu_\infty = A\mu + B\bar{\mu} \tag{13}$$

where $\mu \in \mathbb{R}^k$, $I_{k \times \infty}$ denotes a matrix composed of the first k rows of the infinite identity matrix and $\bar{\mu} \in \mathbb{R}^r$ contains all the moments that appear in the first k elements of $A_\infty \mu_\infty$ but that do not appear in μ.

In this paper we let the vector $\mu \in \mathbb{R}^k$ contain all the moments of $\mathbf{x}$ of order upto $\mathbf{M} \in \mathbb{N}_{\geq 2}$, i.e., we consider an $\mathbf{M}^{th}$ order truncation. With this, the evolution of vector μ can be written as (13) for some matrices A and B with $\bar{\mu} \in \mathbb{R}^r$ being a vector of moments of order $\mathbf{M}+1$. Our goal now is to approximate (13) by a finite-dimensional nonlinear ODE of the form

$$\dot{\nu} = A\nu + B\bar{\varphi}(\nu), \ \nu = [\nu^{(\mathbf{m}1)}, \nu^{(\mathbf{m}2)}, \ldots, \nu^{(\mathbf{m}k)}]^T \tag{14}$$

where the map $\bar{\varphi} : \mathbb{R}^k \to \mathbb{R}^r$ should be chosen so as to keep $\nu(t)$ close to $\mu(t)$. This procedure is commonly referred to as *moment closure*. We call (14) the *truncated moment dynamics* and each element $\varphi^{(\bar{\mathbf{m}})}(\mu)$ of $\bar{\varphi}(\mu)$ the *moment closure function* for the corresponding element $\mu^{(\bar{\mathbf{m}})}$ in $\bar{\mu}$.

The procedure we adopt to construct these moment closure functions is to first assume a certain *separable form* for each element $\varphi^{(\bar{\mathbf{m}})}(\mu)$ of $\bar{\varphi}(\mu)$ and then matching time derivatives of μ and ν at some initial time t_0, for every deterministic initial condition

of the form $\mathbf{x}(t_0) = \bar{\mathbf{x}}$ with probability one. The choice of initial conditions is justified by the fact that the class of deterministic distributions forms a natural basis for the infinite dimensional space Ω_∞ containing every possible state μ_∞ of (11). Referring the reader to [17] for further details the above procedure leads to the following moment closure functions. Assume that each entry $\varphi^{(\bar{\mathbf{m}})}(\mu)$ of $\bar{\varphi}$ has the following separable form

$$\varphi^{(\bar{\mathbf{m}})}(\nu) = \prod_{p=1}^{k} \left(\nu^{(\mathbf{m}p)}\right)^{\gamma_p} = \nu^{(\gamma)}, \;\; \gamma = (\gamma_1, \ldots, \gamma_k) \quad (15)$$

with γ_p are the unique solution of the following system of linear equations[1]

$$\mathrm{C}^{(\bar{\mathbf{m}})}_{(\mathbf{m}s)} = \sum_{p=1}^{k} \gamma_p \mathrm{C}^{(\mathbf{m}p)}_{(\mathbf{m}s)}, \;\; \forall s = \{1, \ldots, k\} \qquad (16)$$

where we define for vectors $\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_n)$ and $\check{\mathbf{m}} = (\check{m}_1, \ldots, \check{m}_n)$

$$\mathrm{C}^{(\hat{\mathbf{m}})}_{(\check{\mathbf{m}})} := \mathrm{C}^{\hat{m}_1}_{\check{m}_1} \mathrm{C}^{\hat{m}_2}_{\check{m}_2} \ldots \mathrm{C}^{\hat{m}_n}_{\check{m}_n}. \qquad (17)$$

Then, when $\mu(t_0) = \nu(t_0)$, we have

$$\frac{d^i \mu(t_0)}{dt^i} = \frac{d^i \nu(t_0)}{dt^i} + \varepsilon_i(\bar{\mathbf{x}}), \;\; \forall i \geq 1$$

for every deterministic initial conditions of the form $\mathbf{x}(t_0) = \bar{\mathbf{x}}$ with probability one for every integer $\bar{\mathbf{x}}$. In the equation above, each entry of the vector $\frac{d^i \mu(t_0)}{dt^i}$ is a polynomial in $\bar{\mathbf{x}}$, whose degree exceeds by at least $\mathbf{M}$ the degree of the corresponding entry of the error vector $\varepsilon_i(\bar{\mathbf{x}})$. Thus, with increasing $\mathbf{M}$, the truncated moment dynamics $\nu(t)$ provides a more accurate approximation to the moments in $\mu(t)$. The striking feature of the moment closure constructed is that they are independent of the reaction parameters (reaction rates and stoichiometry) and moreover the dependence of higher-order moment on lower order ones is consistent with $\mathbf{x}$ being jointly *lognormally distributed*, in spite of the fact that the procedure used to construct φ did not make any assumption on the distribution of the population. For a three-specie reaction ($n = 3$) and a second order truncation $\mathbf{M} = 2$, i.e,

$$\mu = [\mu^{(1,0,0)}, \mu^{(0,1,0)}, \mu^{(0,0,1)}, \mu^{(2,0,0)}, \mu^{(0,2,0)}, \mu^{(0,0,2)}, \mu^{(1,1,0)}, \mu^{(1,0,1)}, \mu^{(0,1,1)}]^T \quad (18)$$

Table 2 lists moment closure functions obtained from (16), which approximate different third order moments of $\mathbf{x}$ in terms of the first two.

Table 2: Moment closure function $\varphi^{(\bar{\mathbf{m}})}(\mu)$ for different third order moments $\mu^{(\bar{\mathbf{m}})}$ with $\mathbf{M} = 2$ and $n = 3$.

$\mu^{\bar{\mathbf{m}}}$	$\varphi^{(\bar{\mathbf{m}})}(\mu)$
$\mu^{(3,0,0)}$	$\left(\frac{\mu^{(2,0,0)}}{\mu^{(1,0,0)}}\right)^3$
$\mu^{(2,1,0)}$	$\left(\frac{\mu^{(2,0,0)}}{\mu^{(0,1,0)}}\right)\left(\frac{\mu^{(1,1,0)}}{\mu^{(1,0,0)}}\right)^2$
$\mu^{(1,1,1)}$	$\frac{\mu^{(1,1,0)}\mu^{(0,1,1)}\mu^{(1,0,1)}}{\mu^{(1,0,0)}\mu^{(0,1,0)}\mu^{(0,0,1)}}$

4 Modeling of Gene Expression with negative feedback

In this section we consider negative feedback in gene expression caused by the protein inhibiting its own transcription. Such auto-regulatory networks are common means of stabilizing protein levels in biochemical pathways [18]. This negative feedback is realized by first assuming that there is an activator that activate a gene to transcribe a mRNA. Proteins are translated from the gene at a constant rate. These proteins can then interact with the activator to change its shape, making it incapable of activation. Hence, more protein implies lesser number of activators, and hence, lesser transcription.

Denoting the mRNA, protein and the activator by X_1, X_2 and X_3, respectively, the above interactions can be written as

$$\begin{aligned}
&\text{Gene} + X_3 \xrightarrow{K_r} \text{Gene} + X_3 + X_1, \\
&X_1 \xrightarrow{d_r} *, \;\; X_1 \xrightarrow{K_p} X_1 + X_2, \;\; X_2 \xrightarrow{d_p} * \\
&* \xrightarrow{K_a} X_3, \;\; X_3 \xrightarrow{d_a} * \\
&X_3 + X_2 \xrightarrow{k_1} X_2 + *.
\end{aligned}$$

[1] C^{ℓ}_{h} is defined as follows: $\forall \ell, h \in \mathbb{N}_{\geq 0}$

$$\mathrm{C}^{\ell}_{h} = \begin{cases} \frac{\ell!}{(\ell-h)!h!}, & \ell \geq h \\ 0, & \ell < h \end{cases}$$

where $\ell!$ denotes the factorial of ℓ.

Hence the mRNA X_1 is transcribed from the gene at a rate $K_r\mathbf{x}_3$, where $\mathbf{x}_3$ denotes the number of molecules of the activator. The protein X_2 is translated from the mRNA at a constant rate K_p. Both mRNA and the protein decay at rates d_r and d_p respectively. K_a and d_a denote the birth and death rate of the specie X_3, respectively. The rate at which the protein X_2 reacts with the activator X_3 and takes it away from the transcription process is given by k_1. Note that increasing values of k_1 denote larger negative feedback. If $k_1 = 0$ then there is no negative feedback. The deterministic chemical rate equations are then given by

$$\dot{\mathbf{x}}_{1_D} = K_r\mathbf{x}_{3_D} - d_r\mathbf{x}_{1_D}, \;\; \dot{\mathbf{x}}_{2_D} = K_p\mathbf{x}_{1_D} - d_p\mathbf{x}_{2_D} \quad (19a)$$

$$\dot{\mathbf{x}}_{3_D} = K_a - (d_a + k_1\mathbf{x}_{2_D})\mathbf{x}_{3_D} \quad (19b)$$

where $\mathbf{x}_{1_D}$, $\mathbf{x}_{2_D}$ and $\mathbf{x}_{3_D}$ are continuous deterministic approximates of $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, respectively. At steady state we have

$$\frac{K_a}{d_a} = \left[1 + \frac{k_1}{d_a}\mathbf{x}_{2_D}(\infty)\right]\mathbf{x}_{3_D}(\infty). \quad (20)$$

As this deterministic model does not provide us with any information about the stochasticity in the protein population, we turn to a stochastic formulation. The above interactions can be modeled as a SHS with reset maps

$$\mathbf{x} \mapsto \phi_1(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 + 1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}, \mathbf{x} \mapsto \phi_2(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 - 1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$$

$$\mathbf{x} \mapsto \phi_3(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 + 1 \\ \mathbf{x}_3 \end{bmatrix}, \mathbf{x} \mapsto \phi_4(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 - 1 \\ \mathbf{x}_3 \end{bmatrix}$$

$$\mathbf{x} \mapsto \phi_5(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 + 1 \end{bmatrix}, \mathbf{x} \mapsto \phi_6(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 - 1 \end{bmatrix}$$

$$\mathbf{x} \mapsto \phi_7(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 - 1 \end{bmatrix}$$

and corresponding transition intensities given by

$$\lambda_1(\mathbf{x}) = K_r\mathbf{x}_3, \;\; \lambda_2(\mathbf{x}) = d_r\mathbf{x}_1, \;\; \lambda_3(\mathbf{x}) = K_p\mathbf{x}_1, \quad (22)$$

$$\lambda_4(\mathbf{x}) = d_p\mathbf{x}_2, \;\; \lambda_5(\mathbf{x}) = K_a, \;\; \lambda_6(\mathbf{x}) = \mathbf{x}_3 d_a \quad (23)$$

$$\lambda_7(\mathbf{x}) = k_1\mathbf{x}_2\mathbf{x}_3. \quad (24)$$

Using (9) the time derivative of the the vector μ which is given by (18) is

$$\dot{\mu} = \bar{\mathbf{A}} + \mathbf{A}\mu + \mathbf{B}\bar{\mu}. \quad (25)$$

for some matrices $\bar{\mathbf{A}}$, $\mathbf{A}$, $\mathbf{B}$ and

$$\bar{\mu} = [\mu^{(0,1,2)}, \mu^{(0,2,1)}, \mu^{(1,1,1)}]^T.$$

Using Table 2 the above system can be closed using the following approximations

$$\mu^{(0,1,2)} = \left(\frac{\mu^{(0,0,2)}}{\mu^{(0,1,0)}}\right)\left(\frac{\mu^{(0,1,1)}}{\mu^{(0,0,1)}}\right)^2 \quad (26a)$$

$$\mu^{(0,2,1)} = \left(\frac{\mu^{(0,0,2)}}{\mu^{(0,0,1)}}\right)\left(\frac{\mu^{(0,1,1)}}{\mu^{(0,1,0)}}\right)^2 \quad (26b)$$

$$\mu^{(1,1,1)} = \frac{\mu^{(1,1,0)}\mu^{(0,1,1)}\mu^{(1,0,1)}}{\mu^{(1,0,0)}\mu^{(0,1,0)}\mu^{(0,0,1)}}. \quad (26c)$$

The statistical variations in the protein population can now be obtained from the closed system (25)-(26). In this paper, we use the steady state *coefficient of variation* defined by

$$CV := \frac{\sqrt{\mathbf{E}[\mathbf{x}_2^2(\infty)] - (\mathbf{E}[\mathbf{x}_2(\infty)])^2}}{\mathbf{E}[\mathbf{x}_2(\infty)]} \quad (27)$$

$$= \frac{\sqrt{\mu^{(0,2,0)}(\infty) - \left(\mu^{(0,1,0)}(\infty)\right)^2}}{\mu^{(0,1,0)}(\infty)} \quad (28)$$

to quantify noise strength in the protein population. As analytical solution of the closed moment equations are too complicated to be of any use we investigate them numerically by taking

$$K_r = 1, \;\; K_p = 5\ sec^{-1}, \;\; d_r = .1\ sec^{-1},$$
$$d_p = .001\ sec^{-1}, \;\; d_a = 100\ sec^{-1}.$$

For comparison purpose, we now vary parameters k_1 and K_a so as to keep $\mathbf{x}_{2_D}(\infty) = 500$ fixed and see its effect on the coefficient of variation. This implies from (20) that $\mathbf{x}_{3_D}(\infty) = .01$ and for a given k_1, K_a is always chosen such that

$$K_a = d_a\left[1 + \frac{k_1}{d_a}\mathbf{x}_{2_D}(\infty)\right]\mathbf{x}_{3_D}(\infty). \quad (29)$$

Table 3 lists the values of CV for different values of k_1 obtained from the steady states of (25)-(26). Note that CV is lower with non-zero values of k_1 than with when there is no feedback, i.e $k_1 = 0$. Hence, we conclude that such negative feedback reduce stochastic variations in the protein numbers.

Table 3: Steady state coefficient of variation in the protein population CV obtained from (25)-(26) for different values of k_1.

k_1	CV
0	.32
.2	.28
.8	.26
1.8	.23

5 Conclusions and Future work

An approximate stochastic model for chemically reacting systems was presented in this paper. This was done by representing the population of various species involved in a set of chemical reactions as the continuous state of a SHS. With such a representation, the dynamics of the infinite vector containing all the statistical moments of the continuous state are governed by an infinite-dimensional linear system of ODEs, which we approximate by finite-dimensional nonlinear ODEs.

This typically involved approximating higher order statistical moments of the population in terms of the lower order moments. Using derivative matching techniques, explicit analytical formulas to construct these approximations were provided. Using then we showed that in gene expression, where a protein inhibits its own transcription, the resulting negative feedback reduces stochastic variations in the protein populations.

An interesting line of future work would be look at gene expression and regulation where the mRNA transcribed from the gene is not immediately accessible for translation. This would corresponds to gene expression in Eukaryotic cells where the mRNA is transported from the nucleus to the cytoplasm where translation takes place. This can be incorporated in our model by introducing another specie that corresponds to the inactive mRNA and is being converted into an active mRNA at some constant rate. Another possible direction of future work would be to investigate stochasticity in gene cascade activation network where protein expressed by one gene activates another gene to express a second protein and negative feedback schemes within this network.

6 Acknowledgements

The authors gratefully acknowledge funding by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office to support this work.

References

[1] M. C. Walters, S. Fiering, J. Eidemiller, W. Magis, M. Groudine, and D. I. K. Martin, "Enhancers increase the probability but not the level of gene expression," *Proceedings of the National Academy of Sciences U.S.A*, vol. 92, pp. 7125–7129, 1995.

[2] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[3] J. L. Spudich and D. E. K. Jr, "Non-genetic individuality: chance in the single cell," *Nature (London)*, vol. 262, pp. 467–471, 1976.

[4] H. H. McAdams and A. P. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of the National Academy of Sciences U.S.A*, vol. 94, pp. 814–819, 1997.

[5] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*. Amsterdam, The Netherlands: Elsevier Science, 2001.

[6] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *J. of Applied Probability*, vol. 4, pp. 413–478, 1967.

[7] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. of Computational Physics*, vol. 22, pp. 403–434, 1976.

[8] D. T. Gillespie and L. R. Petzold, "Improved leap-size selection for accelerated stochastic simulation," *J. of Chemical Physics*, vol. 119, no. 16, pp. 8229–8234, Oct. 2003.

[9] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. of Physical Chemistry A*, vol. 104, pp. 1876–1889, 2000.

[10] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.

[11] J. Paulsson, "Summing up the noise in gene networks," *Nature (London)*, vol. 427, pp. 415–418, 2004.

[12] J. P. Hespanha and A. Singh, "Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems," *Int. J. of Robust and Nonlinear Control*, vol. 15, pp. 669–689, 2005.

[13] A. Singh and J. P. Hespanha, "Models for multi-specie chemical reactions using polynomial stochastic hybrid systems," in *Proc. of the 44th Conf. on Decision and Control, Seville, Spain*, 2005.

[14] ——, "Moment closure techniques for stochastic models in population biology," in *Proc. of the 2006 Amer. Control Conference*, 2006.

[15] F. Wilkinson, *Chemical Kinetics and Reaction Mechanisms*. New York: Van Nostrand Reinhold Co, 1980.

[16] J. P. Hespanha, "Stochastic hybrid systems: Applications to communication networks," in *Hybrid Systems: Computation and Control*, ser. Lect. Notes in Comput. Science, R. Alur and G. J. Pappas, Eds. Berlin: Springer-Verlag, Mar. 2004, no. 2993, pp. 387–401.

[17] A. Singh and J. P. Hespanha, "Lognormal moment closures for biochemical reactions." 2006.

[18] A. Becskei and L. Serrano, "Engineering stability in gene networks by autoregulation," *Nature (London)*, vol. 405, pp. 590–593, 2000.